Title page

Title:

Validating the factor structure and testing measurement invariance of modified Short Form McGill Pain Questionnaire (Ortho- SF-MPQ) for orthodontic pain assessment

Author

Satpal S. Sandhu

PhD Research Postgraduate (Advanced Quantitative Methods), Centre for Multilevel Modelling, Graduate School of Education, University of Bristol, United Kingdom

Formerly Professor and Head, Department of Orthodontics and Dentofacial Orthopedics, Genesis Institute of Dental Sciences and Research, Ferozepur, Punjab, India

Corresponding author: Satpal Singh Sandhu, Helen Wodehouse Building, 35 Berkeley Square, Clifton, Bristol BS8 1JA, United Kingdom. E-mail: drsatpalsandhu@yahoo.co.in

Short Running Title: Validating Short Form McGill Pain Questionnaire

Contributors: SSS was responsible for development and design of the work, as well as the acquisition, analysis and interpretation of data. SSS drafted and revised manuscript and approved the version to be submitted for publication. SSS will act as a guarantor for the paper and accepts full responsibility for the conduct of the study, had access to the data, and controlled the decision to publish.

Funding: No funding

Disclosure of interest: The author report no conflicts of interest

Ethical approval: The study protocol was approved by the ethics committee of the Indian Medical Association, Jalandhar, Punjab, India (April 29, 2013).

Validating the factor structure and testing measurement invariance of modified Short Form McGill Pain Questionnaire (Ortho- SF-MPQ) for orthodontic pain assessment

Abstract

Objective To validate the factor structure of recently modified Short Form McGill Pain Questionnaire (Ortho- SF-MPQ) to assess orthodontic pain**;** and to test its Measurement Invariance (MI) across gender

Methods 180 orthodontic patients were enrolled in this study. 0.016 inch Super-elastic NiTi arch wire was used in 0.022”x0.028” slot pre-adjusted edgewise appliance. After initial arch wire placement, pain was assessed at T1 (24 hours), T2 (day 3), and T3 (day 7) by using the Ortho-SF-MPQ which consists of 7 sensory (pressure, sore, aching, tight, throbbing, pulling, miserable) and 4 affective (uncomfortable, strange, frustrating, annoying) descriptors. Confirmatory factor analysis (CFA) models were fitted for analysis. Multiple-groups CFA (MG-CFA) approach was used for MI testing.

Results Data from 172 patients (85 male, 87 female) with mean age 14.2 years (SD 1.4) was analyzed. CFA model fit indices value at T1 (RMSEA 0.048; CFI 0.995; TLI 0.995), T2 (RMSEA 0.051; CFI 0.998; TLI 0.997), and T3 (RMSEA 0.040; CFI 0.998; TLI 0.998) confirmed the validity of two factor structure of Ortho- SF-MPQ in assessing orthodontic pain. MG-CFA model based non-significant scaled chi-square difference test (Satorra-Bentler method) for *weak invariance* (T1: $\chi2$=6.566, *df*=9, p=0.682; T2: $\chi2$=14.637, *df*=9, p=0.101; T3 ($\chi2$=14.248, *df*=9, p=0.114) and *strong invariance* (T1: $\chi2$=25.874, *df*=20, p=0.170; T2: $\chi2$=25.052, *df*=20, p=0.199; T3: $\chi2$=18.889, *df*=20, p=0.529) confirmed MI across male and female groups.

Conclusion Two factor structure (sensory and affective) of Ortho- SF-MPQ is structurally valid and invariant to measure pain in male and female orthodontic patents after initial arch wire placement.

Key words: Orthodontic pain, Initial arch wire, Ortho- SF-MPQ, Validity, Measurement Invariance

Introduction

Pain is a multidimensional phenomenon characterized by its sensory (location/severity), and affective (generalized well-being/emotional) components (Melzack, 1987). Orthodontic pain affects large number of patients undergoing fixed orthodontic treatment (Bergius *et al.*, 2002) and therefore, is a major concern for patients as well as for orthodontist. Orthodontic pain is characterized by individual variability (e.g. gender based) as well as distinct pattern wherein pain reaches at peak level after 24 hours of force application, start decreasing significantly after 3 days and then declines to baseline level towards the end of one week time period (Bergius *et al.*, 2002, Bergius *et al.*, 2008, Sandhu and Sandhu, 2013b, Sandhu and Sandhu, 2013a).

Although scales like visual analogue scale (VAS), numerical rating scale (NRS) and verbal rating scale (VRS) have been frequently and successfully used in orthodontic pain assessment, these scales record only the intensity of pain sensation and lack the ability to assess the qualitative aspects of the personal experience such as sensory and affective components (Breivik *et al.*, 2008).

Multidimensional assessment of pain by using the MPQ (McGill Pain Questionnaire) or its short form, the SF-MPQ (Short-Form McGill Pain Questionnaire) have become "gold standards" in the measurement of the various qualities of acute and chronic pain. Both forms have been shown to be psychometrically sound, valid, and reliable instruments with good discriminative capacity (Turk and Melzack, 2011).

Recently, Iwasaki et al (Iwasaki *et al.*, 2013) have adapted the Short Form McGill Pain Questionnaire (SF-MPQ) to assess orthodontic pain in adolescents and explored its factor structure by using the exploratory factor analysis (EFA). Authors successfully extracted two factor structure (sensory and affective), as was for the original SF-MPQ (Melzack, 1987).

EFA is generally a descriptive procedure which is typically used earlier in the process of scale development. Once the underlying structure has been tentatively established by using EFA, a more stringent psychometric measurement technique called confirmatory factor analysis (CFA) is used in the later phase to investigate the factor structure of the scale itself and the construct that it purports to assess (Brown, 2015). A key strength of CFA is its ability to test Measurement Invariance (MI) based on the Multiple-groups CFA (MG-CFA) approach. MI determine how well the measurement models generalize to subgroups (e.g., gender) of the population (Brown, 2015).

The objectives of this study were to validate the proposed two factor structure of the recently modified Short Form McGill Pain Questionnaire used for orthodontic pain assessment (Ortho- SF-MPQ) in adolescents (Iwasaki *et al.*, 2013); and to evaluate the MI between male and female groups at three pre-specified time periods i.e. T1 (24 hours), T2 (day 3), and T3 (day 7) after initial arch wire placement. It was decided that if MI is established successfully, then MG-CFA would be continued to test the structural invariance (SI). Unlike MI, which is concerned with the scale validity, SI is not part of validating the scale construct, but rather used to compare subgroups for parameters related to factor variables once the MI has been established (Brown, 2015). In other words, SI is akin to testing the population heterogeneity, and it is normal and expected to have structural non-invariance across subgroups (Brown, 2015).

Methods

Sample size calculation

The details of sample size estimation are provided in the section A of Online Supplementary Material. Power analysis revealed that to achieve 80% power for RMSEA value of 0.05 (good model fit index value) at a two-sided significance level of 0.05 for a CFA model with a *df* 97, a total sample size of 167 participants was required. In this current study, the *df* 97 represents the baseline model for establishing the factor structures across subgroups and the first step in testing the MI.(Brown, 2015) Similarly, using the change in RMSEA values of 0.015 between two nested models (delta RMSEA) to test lack of MI (Brown, 2015), power analysis showed that sample size of 167 participants would provide a power ranging from 85% to 95% for estimation of various levels of MI (depending on the *df* of nested models).

Inclusion criteria were: 1) 11 – 17 year old males and females undergoing full-arch maxillary and mandibular fixed orthodontic treatment, 2) eruption of all maxillary and mandibular teeth except second and/or third molars, 3) moderate to severe crowding, but not severe enough to prevent bracket engagement, 4) no severe deep bite which could affect bracket placement on mandibular anterior teeth or required any treatment other than continuous arch wire for its correction, 5) no history of medical problem/medication which may influence the rate of tooth movement and pain perception, 6) no other intervention including intra-arch or inter-arch elastics, lip bumpers, maxillary expansion appliances etc. required.

Total 180 consecutive patients (90 males, mean age 14.3 years (SD 1.4); 90 females mean age 14.2 years (SD 1.5) who visited the xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx for orthodontic treatment were enrolled in this study after taking written informed consent. The study protocol was approved by the ethics committee of the Indian Medical Association.

The study sample was drawn from an urban population in the North India. All participants were studying in English schools and had good understanding of English language. This was validated during the initial trial run of study wherein a pilot questionnaire was used to: a) assess the understanding of the questionnaire written in English b) to evaluate participant's compliance in reporting the outcome and c) to test the overall feasibility of this study

0.016 inch Super elastic NiTi (austenitic active, Super elastic arch wire; 3M Unitek Corporation, Monrovia, Calif. USA) wires were used in 0.022”x0.028” slot pre-adjusted edgewise appliance (Roth prescription, Gemini Metal Brackets, 3M Unitek Corporation) bonded to maxillary and mandibular dentition using light cure composite resin (Transbond XT, 3M Unitek Corporation). Maxillary and mandibular first molars were banded.

On the day of bonding, patients were provided with questionnaires (written in English), including the written instruction for outcome assessment; and were requested to return the questionnaires after one week either through mail or in-person. The outcome pain was assessed by using the Ortho-SF-MPQ consisting of 7 sensory (pressure, sore, aching, tight, throbbing, pulling, miserable) and 4 affective (uncomfortable, strange, frustrating, annoying) descriptors. Outcome was assessed at three pre-specified time period i.e. T1 (24 hours), T2 (day 3), and T3 (day y) after initial arch wire placement. Patents were asked to rate each of the 11 descriptors on a 4-point Likert response scale (0 = no response, 1 = mild response, 2 = moderate response, and 3 = severe response). A research assistant collected data from the questionnaires returned by the participants. Data entry and transfer of data was double checked for any error by the principal investigator.

Ortho-SF-MPQ also includes a present pain intensity (PPI) scale and 100 mm Visual Analogue Scale (VAS) for assessment of pain. However, since both of these two additional components of Ortho-SF-MPQ have already been validated and collaborated in relation to the orthodontic pain (Iwasaki *et al.*, 2013), the focus of this current study was to analyze the factor structure and MI of Ortho-SF-MPQ. Therefore, the PPI and VAS scores would be presented only as summary statistics.

Statistical analysis

All analyses were performed in R (version 3.2.3) software (R Core Team, 2016). The confirmatory factor analysis (CFA) models were estimated using the Mplus (version 6.12) software (Muthen and Muthen, 2010), calling it from within the R by using the 'MplusAutomation' (version 0.6-3) package (Hallquist and Wiley, 2014). The descriptive statistics were used to describe the score for each individual descriptor. The term 'descriptor' is analogues to 'indicator' in the context of CFA model language.

The models were fitted by using the robust weighted least squares (WLSMV) estimation method which is: a) an appropriate method of estimation for non-normal categorical/ordinal data, b) efficient in handling the missing data, and c) performs well for categorical variables with floor or ceiling effects (Asparouhov and Muthén, 2010, Muthen and Muthen, 2010, Brown, 2015). Ordinal data based omega coefficient (internal consistency) was calculated from the polychoric correlation matrix derived from the ordinal (Likert type) data, as recommended (Gadermann *et al.*, 2012).

The details of model estimation and model identification are provided in the section B1 of Online Supplementary Material. The Section B2 of Online Supplementary Material provides details about the recommended approach used for validating the factor structure and to test the MI based on following steps (Brown, 2015): (1) test the CFA model separately in each group; (2) conduct the simultaneous test of equal form (*configural invariance*); (3) test the equality of factor loadings (*weak/metric invariance)*; (4) test the equality of indicator thresholds (*strong invariance)*; (5) test the equality of factor variances; (6) test the equality of factor co-variances; and (7) test the equality of factor means. The steps 1-4 evaluate the MI whereas steps 5-7 assess the SI. Details of model fit evaluation indices employed at each step is provided in the section B3 of Online Supplementary Material.

Results

Out of total 180 patients enrolled in this study, data obtained from 172 patients (85 male, mean age 14.2 years (SD 1.4); 87 female mean age 14.3 years (SD 1.6) was included in the analysis. Eight participants either did not return the questionnaire or data was missing consistently for one or more variables; therefore, were excluded from the analysis.

For the remaining 172 participants, there was no systematic missing data at any of the three time points, and therefore, were included in the analysis at each time point. Principal investigator randomly cross checked the entered data. Error rate (data entry) was less than 1% and all subsequent corrections were done based on the raw questionnaire data.

The missing data was less than 8.5 % for any variable at any time point. The WLSMV estimation in Mplus software (as in this study) perform estimation with pairwise deletion (unlike the list wise deletion) and therefore, yields unbiased estimation even when data is missing up to 26% for each variable and covariate might have effect on the missing pattern (Asparouhov and Muthén, 2010).

The demographic characteristic data as well as the outcome summary is provided in the Table 1. The highest pain intensity (VAS score and PPI score) reported by both male and female groups was at T1 (24 hrs.). The descriptor score i.e. sum of all eleven descriptors, is also in general agreement with pain intensity scores. The mean and SD, along with the median and quantile (25^{th} and 75^{th}) distribution of each descriptor score is provided in the Table 2. The normality assumption was severely violated, which justifies the fitting of ordinal data based CFA models. The mean scores for sensory and affective descriptors are summarized as Figure S1 and Figure S2 respectively of the Online Supplementary Material. Additional summary statistics data (frequency and percentage of 'yes' response to each individual descriptor) is provided in the Table S1 of Online Supplementary Material.

The omega coefficient (Internal consistency) estimates for both sensory and affective dimension were good to excellent. For male subsample, the coefficient values were 0.899, 0.898, and 0.943 at time T1, T2, and T3 for sensory dimension; and 0.882, 0.959, and 0.961 at time T1, T2, and T3 for affective dimension. For female subsample, the coefficient values were 0.943, 0.899, and 0.944 at time T1, T2, and T3 for sensory dimension; and 0.961, 0.961, and 0.962 at time T1, T2, and T3 for affective dimension.

The results for CFA models are provided in the Table 3 though Table 5. The results show that the two-factor (sensory and affective) models conducted separately for female and male subsamples at each time point were acceptable in terms of all key aspects of the model fit evaluation.

The factor loadings as well as the factor variance and factor co-variances derived from the baseline model (i.e. equal form model) at each time point are shown as Figure 1 through Figure 3; and summarized in Table S2 through Table S4 of Online Supplementary Material. The equal form model (*configural invariance*) provided a good fit to the data at all three time points. All fit indices values were well within the range of good fit at T1 (RMSEA 0.048; CFI 0.995; TLI 0.995), T2 (RMSEA 0.051; CFI 0.998; TLI 0.997), and T3 (RMSEA 0.040; CFI 0.998; TLI 0.998).

The equal factor loadings model (*weak invariance*) fitted data well at each time point. The acceptable model fit indices values at T1 (RMSEA 0.036; CFI 0.997; TLI 0.996), T2 (RMSEA 0.057; CFI 0.997; TLI 0.997), and T3 (RMSEA 0.058; CFI 0.995; TLI 0.995); and nonsignificant scaled chi square test at T1 ($\chi2$ static=6.566, *df*=9, p=0.682), T2 ($\chi2$ static=14.637, *df*=9, p=0.101), and T3 ($\chi2$ static=14.248, *df*=9, p=0.114), as compared to *configural invariance*, established the *weak invariance.*

The equal factor loadings and equal indicator threshold model (*strong invariance*) found to be good-fitting at each time point with the model fit indices values at T1 (RMSEA 0.038; CFI 0.995; TLI 0.996), T2 (RMSEA 0.054; CFI 0.997; TLI 0.997), and T3 (RMSEA 0.050; CFI 0.996; TLI 0.996) falling within the accepted limits. A nonsignificant scaled chi square test at T1 ($\chi2$ static=25.874, *df*=20, p=0.170), T2 ($\chi2$ static=25.052, *df*=20, p=0.199), and T3 ($\chi2$ static=18.889, *df*=20, p=0.529), as compared to *weak invariance*, successfully established the *strong invariance.*

Results for structural invariance (SI) showed a strong evidence for the population heterogeneity. The significant scaled chi square test at T1 ($\chi2$ static=13.556, *df*=2, p=0.001), T2 ($\chi2$ static=6.674, *df*=2, p=0.036), and T3 ($\chi2$ static=8.185, *df*=2, p=0.017) suggests non-equivalent variability of sensory and affective pain across male and female groups. Further, except for time T2, the strength of relationship between the sensory and affective dimensions differed significantly for male and female groups at time T1 ($\chi2$ static=4.159, *df*=1, p=0.041) and T3 ($\chi2$ static=3.859, *df* =1, p=0.049).

Lastly, the significant scaled chi square test at T1 ($\chi2$ static=6.489, *df*=2, p=0.039), T2 ($\chi2$ static=6.618, *df*=2, p=0.037), and T3 ($\chi2$ static=6.118, *df*=2, p=0.047) shows that male and female groups had non-equivalent levels of mean score for sensory and affective dimensions. Compared with male group, the female group showed significantly higher sensory mean score at T1 (0.326, p=0.036) and T2 (0.361, p=0.029). Interestingly, there was no significant gender difference in the mean score for affective dimension at T1 and T2; however, females showed significantly higher mean affective score at T3 (0.345, p=0.041).

Discussion

The purpose of this study was to validate the factor structure and to test the MI (across gender) of recently modified Ortho- SF-MPQ for its intended use to assess orthodontic pain in adolescent population. Unlike previous studies which investigated differences in the orthodontic pain intensity (by using VAS scale, numeric rating scale etc.) amongst male and female subjects undergoing orthodontic treatment, current study is perhaps the first study which evaluated gender differences for orthodontic pain quality (sensory and affective).

The findings of this study confirm the two factor structure of Ortho- SF-MPQ, as proposed by Iwasaki et al (Iwasaki *et al.*, 2013). Hence, two dimensions (sensory and affective) model of the Ortho- SF-MPQ seems to be the most appropriate and informative in assessing orthodontic pain in both male and female adolescent populations.

The results for internal consistency support the fact that Ortho- SF-MPQ is consistent with regards to its internal construct stability. The good to excellent estimate of internal consistency for both sensory and affective dimensions are in agreement with previous study's findings (Iwasaki *et al.*, 2013). However, unlike earlier study which used Cronbach's alpha to estimate the internal consistency, the omega coefficient used in this study is preferred as it outperforms the conventional Cronbach's alpha (Brown, 2015), especially for multidimensional scale, as is the Ortho- SF-MPQ.

The MG-CFA based test of *configural invariance* confirmed the equivalence in the pattern of factor loadings across gender, thereby suggesting that Ortho- SF-MPQ measures the same construct across male and female adolescent orthodontic patients (Brown, 2015). The MG-CFA analysis also established *weak/metric invariance* which implies that eleven descriptors of Ortho- SF-MPQ capture orthodontic pain in a similar way across male and female orthodontic patients (Brown, 2015). Further, the *strong invariance* across two subgroups demonstrates that both male and female orthodontic patients endorsed similar response to each category of threshold of individual indicators. Put it in another way, male and female patients did not differ significantly in terms of their jump from one threshold to another threshold of indicator variables.

Successful establishment of strong MI across male and female subgroups allowed consequent testing for SI. Results showed that at each time point, both male and female patient varied significantly from each other in terms of their response to sensory and affective dimensions of orthodontic pain. These findings are in agreement with the previous studies which claimed that there exists a great between– and within-individual variability in male and female populations with regards to orthodontic pain perception (Bergius *et al.*, 2002, Bergius *et al.*, 2008, Sandhu and Leckie, 2016).

However, interestingly, there was a consistent decrease in the variance of affective dimension for female group across study's time period (Figure 1 through Figure 3). This implies that compared to 24 hours' time period, the response to affective dimensions was becoming more alike and consistent for female patients at day 3 and day 7. This resulted in significant temporal change in the factor co-variance for female group across three time points.

Further, male and female groups showed significant heterogeneity in terms of factor mean scores. Compared to male group, mean sensory score was significantly higher for female group at 24 hours i.e. during the peak pain intensity time. This finding supports the results of previous studies which claimed that females report significantly greater orthodontic pain as compared to male counterparts, especially during the time of peak pain intensity level (Bergius *et al.*, 2002, Bergius *et al.*, 2008).

For the affective dimension, mean score was higher for female group at all three time point, and this difference was statistically significant at T3 (day 7). This finding supports the emerging evidence which suggests that females respond more to the affective/generalized dimension of pain (Rhudy and Williams, 2005, Hood *et al.*, 2013). Evidence suggests that gender differences in the reporting of pain may arise from the differences in the experience and processing of emotion that, in turn, differentially alter pain processing (Rhudy and Williams, 2005). Psychosocial responses to acute pain are possible mechanisms through which these effects occur in females (Hood *et al.*, 2013). These findings have interesting clinical implications for effective management of pain. For instance, evidence shows that gender of an individual may be influential in determining the relative effectiveness of various distraction based strategies for pain management (Thompson *et al.*, 2012). A recent orthodontic study also suggests that the effects of physical activity on reducing pain via enhancement of overall wellbeing of an individual seems gender dependent phenomenon (Sandhu and Sandhu, 2015).

Strengths and limitations

This is, perhaps, the first study which evaluated the factor structure and MI of a well-accepted and widely used multidimensional scale recently modified for orthodontic pain assessment. The successful validation of factor structure and MI across male and female groups would enable orthodontist to use this scale for multidimensional assessment of pain. Matching male and female groups for age, which could otherwise act as a potential confounder, imparts confidence in the validity of study's findings. A step-wise and comprehensive statistical analysis of data ensures that conclusions are reproducible and based on unbiased estimates.

However, this study has few limitations. It is desirable that longitudinal measurement invariance (L-MI) including test-retest reliability across time should be evaluated once cross-sectional MI is established. The reason for not proceeding with L-MI testing in this study was inadequate sample size required for L-MI because the number of parameters increases substantial in longitudinal CFA framework. Another limitation pertains to the fact that only one week's time period was considered for MI testing. It is quite possible that the result based on longer time period may differ from this study. Therefore, longer time period of MI evaluation are warranted in future studies. Lastly, age based MI was not performed in this study owing to the sample size constraint. This because including both age and gender, and their interaction effect, could have adversely affected the statistical power.

It is recommended that future studies based on a larger sample size should extend the work presented in this study by exploring a combined influence of age and gender (as well as their interaction effect) on the factor structure as well as the MI over a longer period of time by undertaking a longitudinal CFA.

Conclusions

1. Two factor structure (sensory and affective) of Ortho- SF-MPQ is valid for orthodontic pain assessment in an adolescent population.
2. The successful establishment of measurement invariance for Ortho- SF-MPQ ensures that the constructs are operationalized similarly across male and female subpopulations.
3. The results of structural invariance showed significant between- and within individual variability of pain perception. Compared to males, female group showed significantly higher sensory pain perception and responded more consistently and strongly to the affective dimension.

References

Asparouhov, T. & Muthén, B., 2010. Weighted least squares estimation with missing data. *Mplus Technical Appendix*, 1-10.

Bergius, M., Berggren, U. & Kiliaridis, S., 2002. Experience of pain during an orthodontic procedure. *European journal of oral sciences,* 110, 92-8.

Bergius, M., Broberg, A.G., Hakeberg, M. & Berggren, U., 2008. Prediction of prolonged pain experiences during orthodontic treatment. *Am J Orthod Dentofacial Orthop,* 133, 339.e1-8.

Breivik, H., Borchgrevink, P.C., Allen, S.M., Rosseland, L.A., Romundstad, L., Hals, E.K.B., Kvarstein, G. & Stubhaug, A., 2008. Assessment of pain. *British journal of anaesthesia,* 101, 17-24.

Brown, T.A., 2015. *Confirmatory Factor Analysis for Applied Research, Second Edition*: Guilford Publications.

Gadermann, A.M., Guhn, M. & Zumbo, B.D., 2012. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation,* 17, 1-13.

Hallquist, M. & Wiley, J., 2014. MplusAutomation: Automating Mplus Model Estimation and Interpretation. R package version 0.6-3.

Hood, A., Pulvers, K. & Spady, T.J., 2013. Timing and Gender Determine if Acute Pain Impairs Working Memory Performance. *The journal of pain,* 14, 10.1016/j.jpain.2013.05.015.

Iwasaki, L.R., Freytag, L.E., Schumacher, C.A., Walker, M.P. & Williams, K.B., 2013. Validation of a modified McGill Pain Questionnaire for orthodontic patients. *The Angle Orthodontist*.

Melzack, R., 1987. The short-form McGill Pain Questionnaire. *Pain,* 30, 191-7.

Muthen, L.K. & Muthen, B.O., 2010. Mplus user's guide. 6th. *Los Angeles, CA: Muthén & Muthén*.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing. R version 3.2.4.

Rhudy, J.L. & Williams, A.E., 2005. Gender differences in pain: do emotions play a role? *Gend Med,* 2, 208-26.

Sandhu, S.S. & Leckie, G., 2016. Orthodontic pain trajectories in adolescents: Exploring the between- and within-subject variability in pain perception. *Am J Orthod Dentofacial Orthop,* 149, 491-500.

Sandhu, S.S. & Sandhu, J., 2013a. Orthodontic pain: an interaction between age and sex in early and middle adolescence. *Angle Orthod,* 83, 966-72.

Sandhu, S.S. & Sandhu, J., 2013b. A randomized clinical trial investigating pain associated with superelastic nickel–titanium and multistranded stainless steel archwires during the initial leveling and aligning phase of orthodontic treatment. *Journal of Orthodontics,* 40, 276-285.

Sandhu, S.S. & Sandhu, J., 2015. Effect of physical activity level on orthodontic pain perception and analgesic consumption in adolescents. *American Journal of Orthodontics and Dentofacial Orthopedics,* 148, 618-627.

Thompson, T., Keogh, E., Chen, M.J.L. & French, C.C., 2012. Emotion-focused coping and distraction: Sex differences in the influence of anxiety sensitivity during noxious heat stimulation. *European Journal of Pain,* 16, 410-420.

Turk, D.C. & Melzack, R., 2011. *Handbook of pain assessment,* 3rd ed. New York: Guilford Press.

Tables

Table 1. Demographics and summary characteristics data. #

	T1 (24 hrs)	T2 (day 3)	T3 (day 7)
Participants, (number, (per cent)			
Male	85 (49.42 %)	85 (49.42 %)	85 (49.42 %)
Female	87 (50.58 %)	87 (50.58 %)	87 (50.58 %)
Overall	172 (100 %)	172 (100 %)	172 (100 %)
Age in years, mean (SD)			
Male	14.2 (1.4)	14.2 (1.4)	14.2 (1.4)
Female	14.3 (1.6)	14.3 (1.6)	14.3 (1.6)
Overall	14.25 (1.5)	14.25 (1.5)	14.25 (1.5)
VAS score, mean (SD)			
Male	33.41 (14.98)	20.07 (16.45)	10.39 (9.32)
Female	40.42 (23.64)	27.02 (21.57)	15.87 (13.59)
Overall	36.95 (20.09)	23.59 (20.33)	13.16 (11.96)
PPI score, median (q1, q3)			
Male	2 (1, 3)	1 (0, 3)	1 (0, 2)
Female	3 (2, 4)	2 (1, 4)	2 (1, 3)
Overall	2 (1, 4)	2 (0, 3)	1 (0, 2)
Descriptors score, median (q1, q3)			
Male	12 (8,15)	7 (5,12)	5 (3,8)
Female	14 (9,17)	10 (7,14)	6.5 (5,10)
Overall	12.5 (9.5,16)	8.5 (6,11.5)	6 (5, 9.5)

mean: Mean; SD: Standard Deviation of the mean; median: Median; q1: 25th quantile; q3: 75th quantile. The Inter-quantile range is q3-q1 and Semi-interquartile range is half the Inter-quantile range i.e. q3-q1 / 2

VAS score: Visual Analogue Scale score in millimetres (range 0 - 100)

PPI score: Present Pain Index score (range 0 - 5)

Descriptors score: Sum of 11 (7 sensory and 4 affective) descriptors (range 0 - 33)

Table 2 Descriptive statistics for each individual descriptor of sensory and affective dimension.*

	Male (N=85)								Female (N=87)							
	median	q1	q3	Mean	SD	Skew	Kurt	p value	median	q1	q3	Mean	SD	Skew	Kurt	p value
T1 (24 hours)																
Sensory																
T1pressure	1	1	2	1.671	0.822	0.405	-1.010	0.000	2	1	3	1.874	0.887	0.145	-1.532	0.000
T1sore	1	1	2	1.565	0.808	0.661	-0.762	0.000	1	1	2	1.690	0.853	0.405	-1.153	0.000
T1aching	1	1	2	1.588	0.863	0.224	-0.816	0.000	2	1	3	1.805	0.887	-0.011	-1.091	0.000
T1tight	1	1	2	1.518	0.881	0.257	-0.773	0.000	1	1	3	1.598	1.017	0.197	-1.244	0.000
T1throbbing	1	0	1	0.965	0.981	0.742	-0.503	0.000	1	0	2	1.287	1.077	0.302	-1.198	0.000
T1pulling	1	1	2	1.529	0.946	0.042	-0.950	0.000	2	1	2	1.690	0.980	-0.165	-1.036	0.000
T1miserable	1	0	1	0.941	1.028	0.831	-0.498	0.000	2	0	3	1.529	1.170	-0.025	-1.494	0.000
Affective																
T1uncomfortable	1	1	2	1.329	1.016	0.260	-1.061	0.000	1	1	2	1.460	0.974	0.112	-1.008	0.000
T1strange	1	0	2	1.012	0.893	0.572	-0.460	0.000	1	0	2	1.172	1.037	0.400	-1.052	0.000
T1frustrating	1	0	2	1.271	0.956	0.094	-1.067	0.000	1	1	2	1.414	1.052	0.077	-1.225	0.000
T1annoying	1	0	2	1.118	0.865	0.322	-0.677	0.000	1	0	2	1.138	0.979	0.462	-0.821	0.000
T2 (day 3)																
Sensory																
T2pressure	1	0	2	0.894	0.976	0.664	-0.800	0.000	1	0	2	1.092	0.996	0.168	-1.405	0.000
T2sore	0	0	1	0.741	0.915	1.076	0.223	0.000	1	0	2	0.966	0.994	0.490	-1.087	0.000
T2aching	1	0	2	0.929	0.923	0.586	-0.717	0.000	1	0	2	1.115	0.920	0.129	-1.203	0.000
T2tight	1	0	1	0.729	0.836	0.892	-0.050	0.000	1	0	1	0.897	0.850	0.532	-0.645	0.000
T2throbbing	0	0	1	0.482	0.811	1.579	1.553	0.000	1	0	2	0.931	0.938	0.470	-1.049	0.000
T2pulling	0	0	1	0.682	0.876	1.176	0.568	0.000	1	0	2	0.931	0.912	0.588	-0.666	0.000
T2miserable	0	0	1	0.459	0.853	1.719	1.794	0.000	1	0	2	0.943	0.969	0.568	-0.890	0.000
Affective																
T2uncomfortable	0	0	2	0.882	1.138	0.802	-0.951	0.000	1	0	2	1.011	1.040	0.652	-0.817	0.000
T2strange	0	0	1	0.741	0.941	1.035	-0.031	0.000	0	0	1	0.736	0.933	1.131	0.298	0.000
T2frustrating	0	0	1	0.694	0.988	1.142	-0.017	0.000	0	0	1	0.793	1.002	0.965	-0.334	0.000
T2annoying	0	0	1	0.553	0.824	1.348	0.902	0.000	0	0	2	0.828	0.979	0.785	-0.667	0.000
T3 (day 7)																
Sensory																
T3pressure	1	0	2	0.800	0.870	0.497	-1.238	0.000	1	0	2	0.977	0.902	0.138	-1.565	0.000
T3sore	0	0	1	0.682	0.790	0.756	-0.564	0.000	1	0	2	0.851	0.883	0.390	-1.375	0.000
T3aching	1	0	2	0.847	0.838	0.407	-1.161	0.000	1	0	2	1.011	0.896	0.074	-1.558	0.000
T3tight	1	0	1	0.871	0.842	0.479	-0.858	0.000	1	0	2	1.069	0.912	0.320	-0.941	0.000
T3throbbing	0	0	1	0.447	0.748	1.443	0.921	0.000	0	0	2	0.701	0.891	0.707	-1.124	0.000
T3pulling	1	0	1	0.812	0.779	0.482	-0.777	0.000	1	0	2	1.011	0.814	0.107	-1.201	0.000
T3miserable	0	0	0	0.341	0.733	1.897	2.212	0.000	0	0	2	0.782	0.882	0.532	-1.262	0.000
Affective																
T3uncomfortable	1	0	2	0.894	0.976	0.740	-0.611	0.000	1	0	2	1.011	0.896	0.170	-1.349	0.000
T3strange	0	0	1	0.588	0.821	1.380	1.298	0.000	0	0	1	0.678	0.869	1.079	0.221	0.000
T3frustrating	0	0	1	0.694	0.964	1.025	-0.316	0.000	1	0	1	0.805	0.874	0.795	-0.282	0.000
T3annoying	0	0	1	0.612	0.832	1.056	-0.038	0.000	0	0	2	0.782	0.908	0.620	-1.074	0.000

* median: Median; q1: 25th quantile; q3: 75th quantile; mean: Mean; SD: Standard Deviation of the mean; Skew: Skewness; Kurt: Kurtosis ; p value: Normality test based probability of deviation from the normal distribution

Table 3 Tests of Measurement Invariance and Structural Invariance (population heterogeneity) for two-factor model of Ortho-SF-MPQ across male and female groups at T1 (24 hours)

Model	Degree of freedom	RMSEA	RMSEA (Lower)	RMSEA (Upper)	RMSEA pvalue	CFI	TLI	Delta RMSEA	Delta CFI	Delta TLI	Chi.Squre DiffTest	Chi.Squre DiffTest DF	Chi.Squre DiffTest pvalue
Single-group solutions													
Female	43	0.040	0.000	0.086	0.591	0.997	0.997	N/A	N/A	N/A	N/A	N/A	N/A
Male	43	0.057	0.013	0.098	0.223	0.985	0.980	N/A	N/A	N/A	N/A	N/A	N/A
Measurement Invariance													
Configural	97	0.048	0.000	0.077	0.528	0.995	0.995	N/A	N/A	N/A	N/A	N/A	N/A
Weak	106	0.036	0.000	0.069	0.721	0.997	0.996	0.012	0.002	0.001	6.566	9	0.682
Strong	126	0.038	0.000	0.067	0.715	0.995	0.996	0.002	0.002	0.000	25.874	20	0.170
Structural Invariance													
Factor variance	128	0.064	0.037	0.087	0.173	0.987	0.989	0.026	0.008	0.007	13.556	2	0.001
Factor covarinace	129	0.069	0.044	0.092	0.096	0.985	0.987	0.005	0.002	0.002	4.159	1	0.041
Factor mean	131	0.074	0.050	0.095	0.050	0.983	0.985	0.005	0.002	0.002	6.489	2	0.039

RMSEA = root-mean-square error of approximation; RMSEA (Lower) = Lower bound of 90% Confidence Interval for RMSEA; RMSEA (Upper) = Upper bound of 90% Confidence Interval for RMSEA; RMSEA pvalue = p value significance level for RMSEA; CFI = comparative fit index; TLI = Tucker–Lewis index; Delta RMSEA = change in RMSEA for nested model; Delta CFI = change in CFI for nested model; Delta TLI = change in TLI for nested model; Chi.Squre DiffTest = Scaled (Satorra - Bentler) chi-square difference test statistic; Chi.Squre DiffTest DF = degree of freedom for Scaled (Satorra - Bentler) chi-square difference test; Chi.Squre DiffTest pvalue = p value significance level for Scaled (Satorra - Bentler) chi-square difference test

Table 4 Tests of Measurement Invariance and Structural Invariance (population heterogeneity) for two-factor model of Ortho-SF-MPQ across male and female groups at T2 (day 3)

Model	Degree of freedom	RMSEA	RMSEA (Lower)	RMSEA (Upper)	RMSEA pvalue	CFI	TLI	Delta RMSEA	Delta CFI	Delta TLI	Chi.Squre DiffTest	Chi.Squre DiffTest DF	Chi.Squre DiffTest pvalue
Single-group solutions													
Female	43	0.005	0.000	0.073	0.793	1.000	1.000	N/A	N/A	N/A	N/A	N/A	N/A
Male	43	0.043	0.000	0.088	0.555	0.998	0.997	N/A	N/A	N/A	N/A	N/A	N/A
Measurement Invariance													
Configural	97	0.051	0.000	0.080	0.471	0.998	0.997	N/A	N/A	N/A	N/A	N/A	N/A
Weak	106	0.057	0.020	0.084	0.333	0.997	0.997	0.006	0.001	0.000	14.637	9	0.101
Strong	126	0.054	0.018	0.079	0.398	0.997	0.997	0.003	0.000	0.000	25.052	20	0.199
Structural Invariance													
Factor variance	128	0.059	0.029	0.083	0.271	0.996	0.996	0.005	0.001	0.001	6.674	2	0.036
Factor covarinace	129	0.046	0.000	0.073	0.565	0.997	0.998	0.013	0.001	0.002	1.245	1	0.265
Factor mean	131	0.057	0.025	0.081	0.330	0.996	0.997	0.011	0.001	0.001	6.618	2	0.037

RMSEA = root-mean-square error of approximation; RMSEA (Lower) = Lower bound of 90% Confidence Interval for RMSEA; RMSEA (Upper) = Upper bound of 90% Confidence Interval for RMSEA; RMSEA pvalue = p value significance level for RMSEA; CFI = comparative fit index; TLI = Tucker–Lewis index; Delta RMSEA = change in RMSEA for nested model; Delta CFI = change in CFI for nested model; Delta TLI = change in TLI for nested model; Chi.Squre DiffTest = Scaled (Satorra - Bentler) chi-square difference test statistic; Chi.Squre DiffTest DF = degree of freedom for Scaled (Satorra - Bentler) chi-square difference test; Chi.Squre DiffTest pvalue = p value significance level for Scaled (Satorra - Bentler) chi-square difference test

Table 5 Tests of Measurement Invariance and Structural Invariance (population heterogeneity) for two-factor model of Ortho-SF-MPQ across male and female groups at T3 (day 7)													
Model	Degree of freedom	RMSEA	RMSEA (Lower)	RMSEA (Upper)	RMSEA pvalue	CFI	TLI	Delta RMSEA	Delta CFI	Delta TLI	Chi.Squre DiffTest	Chi.Squre DiffTest DF	Chi.Squre DiffTest pvalue
Single-group solutions													
Female	43	0.000	0.000	0.050	0.948	1.000	1.002	N/A	N/A	N/A	N/A	N/A	N/A
Male	43	0.056	0.000	0.097	0.397	0.994	0.993	N/A	N/A	N/A	N/A	N/A	N/A
Measurement Invariance													
Configural	97	0.040	0.000	0.072	0.661	0.998	0.998	N/A	N/A	N/A	N/A	N/A	N/A
Weak	106	0.058	0.023	0.084	0.312	0.997	0.996	0.018	0.001	0.002	14.248	9	0.114
Strong	126	0.050	0.000	0.076	0.487	0.996	0.996	0.008	0.001	0.000	18.889	20	0.529
Structural Invariance													
Factor variance	128	0.060	0.031	0.084	0.246	0.994	0.995	0.010	0.002	0.001	8.185	2	0.017
Factor covarinace	129	0.059	0.028	0.083	0.281	0.994	0.995	0.001	0.000	0.000	3.859	1	0.049
Factor mean	131	0.064	0.037	0.087	0.173	0.993	0.994	0.005	0.001	0.001	6.118	2	0.047
RMSEA = root-mean-square error of approximation; RMSEA (Lower) = Lower bound of 90% Confidence Interval for RMSEA; RMSEA (Upper) = Upper bound of 90% Confidence Interval for RMSEA; RMSEA pvalue = p value significance level for RMSEA; CFI = comparative fit index; TLI = Tucker–Lewis index; Delta RMSEA = change in RMSEA for nested model; Delta CFI = change in CFI for nested model; Delta TLI = change in TLI for nested model; Chi.Squre DiffTest = Scaled (Satorra - Bentler) chi-square difference test statistic; Chi.Squre DiffTest DF = degree of freedom for Scaled (Satorra - Bentler) chi-square difference test; Chi.Squre DiffTest pvalue = p value significance level for Scaled (Satorra - Bentler) chi-square difference test													

Figures

Figure 1 Baseline model (equal form) solutions at T1 (24 hours)

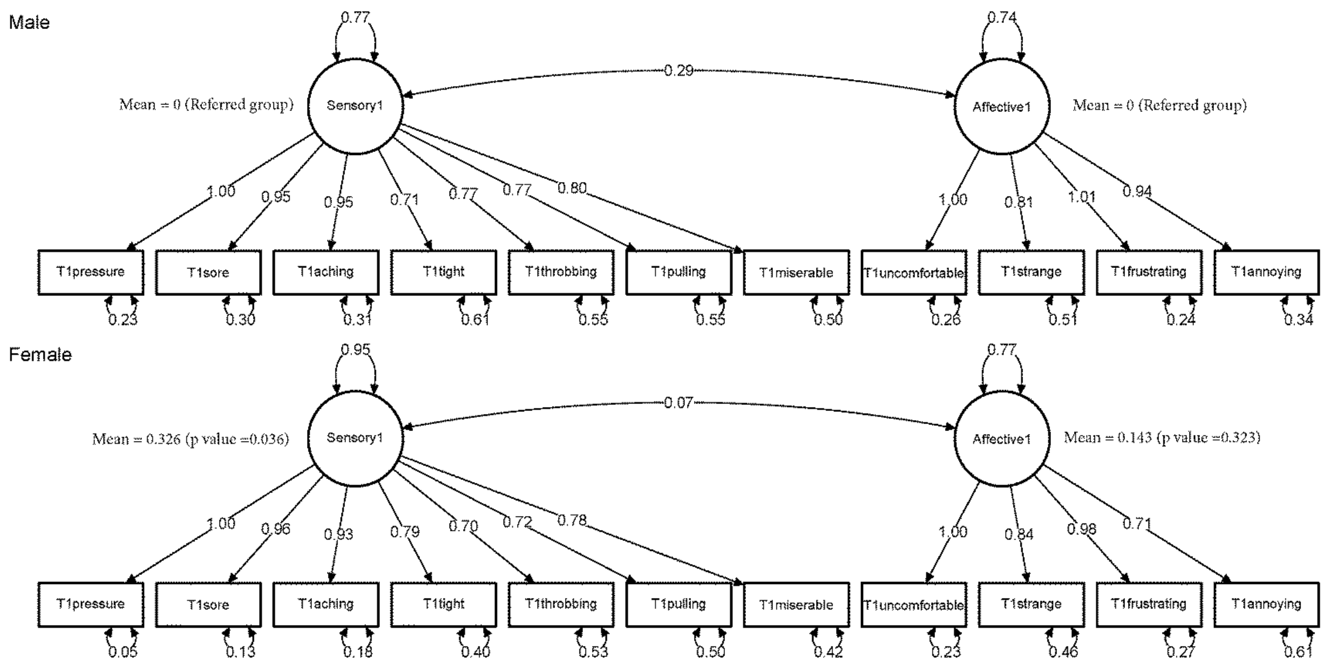


Figure 2 Baseline model (equal form) solutions at T2 (day 3)

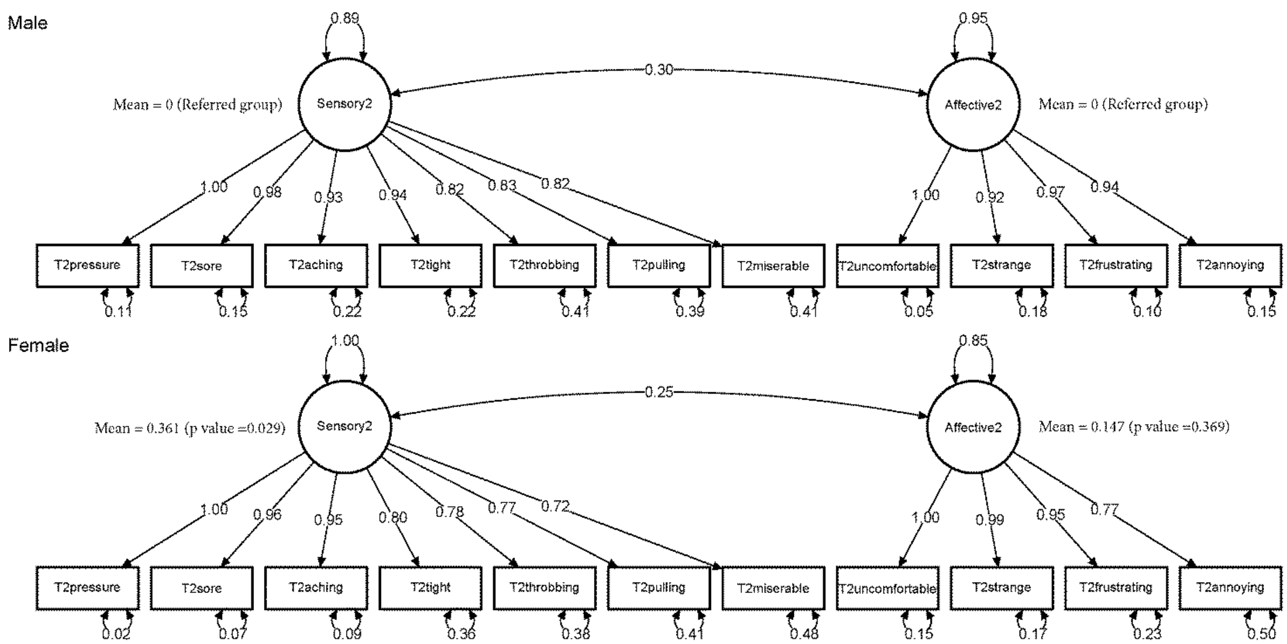
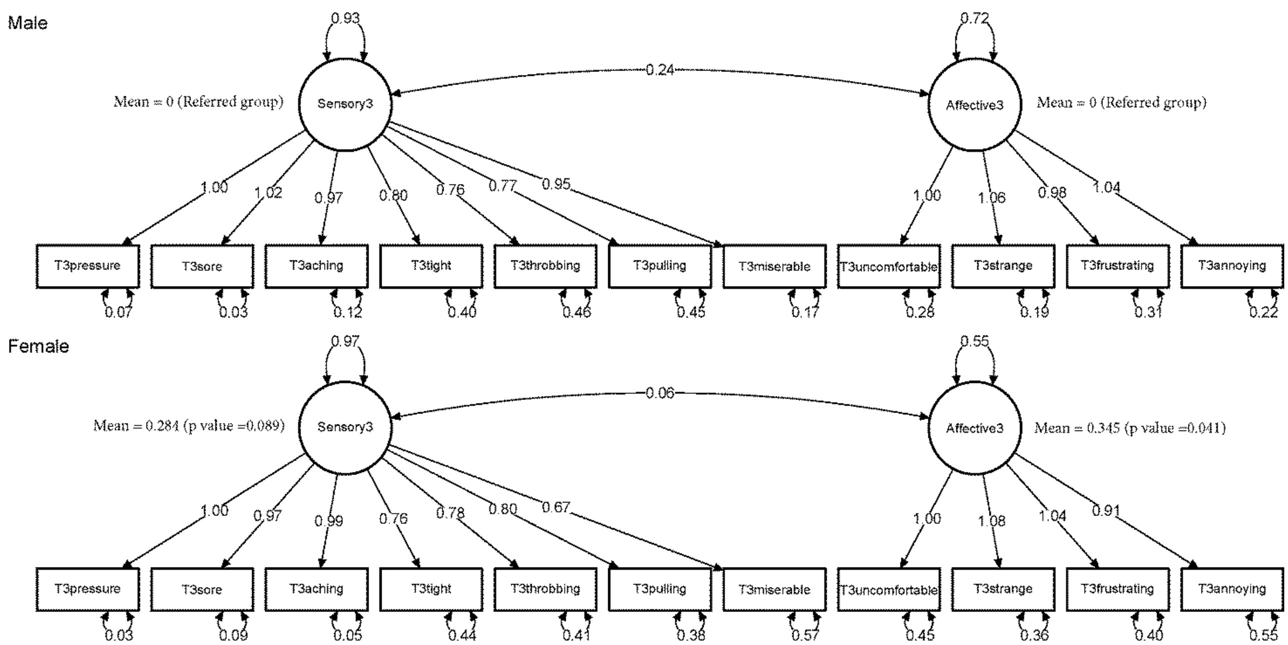


Figure 3 Baseline model (equal form) solutions at T3 (day 7)



Tables

Table 1 Descriptive statistics for each individual descriptor

Table 2 Summary of Measurement Invariance testing at T1 (24 hours)

Table 3 Summary of Measurement Invariance testing at T2 (day 3)

Table 4 Summary of Measurement Invariance testing at T3 (day 7)

Figure captions

Figure 1 Baseline model (equal form) solutions at T1 (24 hours)

Figure 2 Baseline model (equal form) solutions at T2 (day 3)

Figure 3 Baseline model (equal form) solutions at T3 (day 7)